

Modifications of Wald's score tests on large dimensional covariance matrices structure.

Dandan Jiang^{a,1,*}, Qibin Zhang^a

^a*School of Mathematics,
Jilin University,
2699 QianJin Street,
Changchun 130012, China.*

Abstract

This paper considers testing the covariance matrices structure based on Wald's score test in large dimensional setting. The hypothesis $H_0 : \Sigma = \Sigma_0$ for a given matrix Σ_0 , which covers the identity hypothesis test and sphericity hypothesis test as the special cases, is reviewed by the generalized CLT (Central Limit Theorem) for the linear spectral statistics of large dimensional sample covariance matrices from Jiang (2015). The proposed tests can be applicable for large dimensional non-Gaussian variables in a wider range. Furthermore, the simulation study is provided to compare the proposed tests with other large dimensional covariance matrix tests for evaluation of their performances. As seen from the simulation results, our proposed tests are feasible for large dimensional data without restriction of population distribution and provide the accurate and steady empirical sizes, which are almost around the nominal size.

Keywords: Large dimensional data, Covariance structure tests, Wald's score test, Random matrix theory

2010 MSC: 62H15, 62H10

*Corresponding author:Dandan Jiang

Email addresses: jiangdandan@jlu.edu.cn (Dandan Jiang),
jasonzhangqb@gmail.com (Qibin Zhang)

¹Supported by Project 11471140 from NSFC.

1. Introduction

The hypothesis of testing the covariance matrices structure is represented as

$$H_0 : \Sigma = \Sigma_0 \quad \text{v.s.} \quad H_1 : \Sigma \neq \Sigma_0, \quad (1.1)$$

which plays an important role in classical multivariate analysis, and it has been widely used in the social and behavioural sciences etc. However, the rapid development of data acquisition and computer science have challenged the traditional statistical methods, because they are established on the basis of fixed dimension p and fail to analyze large dimensional data. This inspired growing interests in proposing new testing procedures designed for the large dimensional data. In this aspect, Johnstone (2001), Ledoit and Wolf (2002) and Srivastava (2005) gave the large dimensional tests based on some different functions of the distance between the null and alternative hypotheses under the Gaussian assumption. Bai et al.(2009) and Jiang et al.(2012) derived the asymptotic distributions of their proposed test statistics in large limiting scheme $p/n \rightarrow c \in [0, 1)$ and $p/n \rightarrow c \in [0, 1]$ with $p < n$, respectively. Recently, Chen et al.(2010) proposed a nonparametric test regardless of the limiting behavior of p/n and Cai and Ma(2013) presented an optimal test on high dimensional covariance matrix from a minimax point of view. Also, Jiang(2015) used the RMT (Random Matrix Theory) to correct the Rao's score test, which is applicable for large dimensional data without restriction of Gaussian assumption.

Motivated by the above works, we reviewed the testing problem (1.1) and proposed some new tests based on amending the Wald's score tests by RMT. We first derived Wald's score test statistic for the hypothesis (1.1) and refined it to satisfy the enhanced version of the CLT(Central Limit Theorem) for LSS (Linear Spectral Statistics) of large dimensional sample covariance matrices in Jiang(2015), which is the basic tools used in the amendment process. Then we proposed the correction to Wald's score test statistics and derived their asymptotic distribution under large limiting scheme $p/(n-1) \rightarrow q \in [0, 1)$, which provided more accurate and steady empirical sizes.

The remainder of the article is arranged as below. A quick survey of Wald's score test and an enhanced version of the large dimensional CLT cited from Jiang(2015) are reviewed in Section 2, and the details of Wald's score tests on covariance structure are also derived in this part. In Section 3, the new testing statistics based on the aforementioned Wald's score tests are proposed by the large dimensional limiting tools in RMT, which compensate

for the effects of large dimensionality. Then the simulations are conducted to evaluate the performance of our proposed tests compared with other large dimensional tests in Section 4. Finally, conclusions and comments are drawn in the Section 5, and the proofs and derivations are listed in the Appendix A.

2. Preliminary

Let $\chi = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ denote a sample from a random vector \mathbf{X} following the population distribution $F_{\mathbf{X}}(x, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown parameter. Make some notations as below:

- $f_{\mathbf{X}}(x, \boldsymbol{\theta})$ is the density function of \mathbf{X} ;
- $U(\mathbf{X}, \boldsymbol{\theta}) = \frac{d}{d\boldsymbol{\theta}} \ln f_{\mathbf{X}}(x, \boldsymbol{\theta})$ is the score vector of \mathbf{X} ;
- $I(\mathbf{X}, \boldsymbol{\theta}) = E(U(\mathbf{X}, \boldsymbol{\theta})U'(\mathbf{X}, \boldsymbol{\theta}))$ is the information matrix of \mathbf{X} , which is also calculated by Hessian matrix $H(\mathbf{X}, \boldsymbol{\theta})$ as below:

$$I(\mathbf{X}, \boldsymbol{\theta}) = -E(H(\mathbf{X}, \boldsymbol{\theta})) = -E\left(\frac{d^2}{d\boldsymbol{\theta}^2} \ln f_{\mathbf{X}}(x, \boldsymbol{\theta})\right)$$

Then the log-likelihood, the score function and the information matrix of the sample are given by $l(\chi, \boldsymbol{\theta}) = \sum_{i=1}^n \ln f(\mathbf{x}_i, \boldsymbol{\theta})$, $U(\chi, \boldsymbol{\theta}) = \sum_{i=1}^n U(\mathbf{x}_i, \boldsymbol{\theta})$ and $I(\chi, \boldsymbol{\theta}) = nI(\mathbf{x}_1, \boldsymbol{\theta})$, respectively. So the definition of Wald's score test statistic is described as below:

Definition 2.1. For hypothesis test $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, Wald's score test statistic(WST) is defined as

$$WST(\chi, \boldsymbol{\theta}_0) = (T_{\boldsymbol{\theta}}(\chi) - \boldsymbol{\theta}_0)' \cdot I(\chi, T_{\boldsymbol{\theta}}(\chi)) \cdot (T_{\boldsymbol{\theta}}(\chi) - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_0 = (\theta_{01}, \dots, \theta_{0p})'$ is a known vector, χ is a random sample from the population distribution, $T_{\boldsymbol{\theta}}(\chi)$ is the maximum likelihood estimator of the parameter $\boldsymbol{\theta}$ and $I(\chi, T_{\boldsymbol{\theta}}(\chi))$ is the information matrix substituting $\boldsymbol{\theta}$ with $T_{\boldsymbol{\theta}}(\chi)$. Then $WST(\chi, \boldsymbol{\theta}_0)$ tends to a χ_p^2 limiting distribution, which is a χ^2 -distribution with p degrees of freedom as $n \rightarrow \infty$ under H_0 . (Wald, 1943).

To figure out the Wald's score test statistic for the hypothesis test (1.1), we suppose that the sample $\chi = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ follows a normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Denote $\boldsymbol{\theta} = (\boldsymbol{\mu}', \text{vec}'(\boldsymbol{\Sigma}))'$, where $\text{vec}(\cdot)$ is the vectorization operator. Then the logarithm of the density of the sample χ is

$$l(\chi, \boldsymbol{\theta}) = -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n \text{tr}(\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})').$$

By some derivations we arrive at the following several results [for details see the Appendix A.1 in Jiang (2015)],

$$U(\chi, \boldsymbol{\theta}) =: \begin{pmatrix} U_1(\chi, \boldsymbol{\theta}) \\ U_2(\chi, \boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} n\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\ \frac{n}{2} \text{vec}(\boldsymbol{\Sigma}^{-1}(\mathbf{S}\boldsymbol{\Sigma}^{-1} - \mathbf{I}_p)) \end{pmatrix} \quad (2.1)$$

where $\frac{d}{d\boldsymbol{\theta}} = \begin{pmatrix} \frac{d}{d\boldsymbol{\mu}} \\ \frac{d}{d\text{vec}(\boldsymbol{\Sigma})} \end{pmatrix}$ is a $p(p+1) \times 1$ vector and

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad \mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})'.$$

According to the definition of Hessian matrix, $H(\chi, \boldsymbol{\theta}) = \frac{d^2}{d\boldsymbol{\theta}^2} l(\chi, \boldsymbol{\theta}) =: \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$, where H_{22} is the part for the parameter $\boldsymbol{\Sigma}$ and written as

$$\begin{aligned} H_{22} &= \frac{n \text{dvec}(\boldsymbol{\Sigma}^{-1}(\mathbf{S}\boldsymbol{\Sigma}^{-1} - \mathbf{I}_p))}{2 \text{dvec}'(\boldsymbol{\Sigma})} \\ &= \frac{n \text{dvec}(\boldsymbol{\Sigma}^{-1}) \text{dvec}(\boldsymbol{\Sigma}^{-1}\mathbf{S}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1})}{2 \text{dvec}'(\boldsymbol{\Sigma}) \text{dvec}'(\boldsymbol{\Sigma}^{-1})} \\ &= -\frac{n}{2} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{S}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_p + \mathbf{I}_p \otimes \mathbf{S}\boldsymbol{\Sigma}^{-1} - \mathbf{I}_{p^2}). \end{aligned} \quad (2.2)$$

where \otimes is the Kronecker products. The information matrix is denoted as $I(\chi, \boldsymbol{\theta}) =: \begin{pmatrix} I_{11}(\chi, \boldsymbol{\theta}) & I_{12}(\chi, \boldsymbol{\theta}) \\ I_{21}(\chi, \boldsymbol{\theta}) & I_{22}(\chi, \boldsymbol{\theta}) \end{pmatrix}$, then we get the part of information matrix for $\boldsymbol{\Sigma}$

$$I_{22}(\chi, \boldsymbol{\theta}) = -E[H_{22}(\chi, \boldsymbol{\theta})] = \frac{n}{2} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})$$

by the expectation $E(\mathbf{S}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}$.

As stated in Gombay (2002), if there are no restrictions on $\boldsymbol{\mu}$, the parameter $\boldsymbol{\mu}$ and \mathbf{S} in the score vector are replaced by its maximum likelihood estimator $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, where

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})'. \quad (2.3)$$

Then we have $U_1(\chi, \boldsymbol{\theta}) = \mathbf{0}$, so only $I_{22}(\chi, \boldsymbol{\theta})$ is involved in the calculation of the Wald's score test statistic. Therefore, we have

Proposition 2.1. *Wald's score test statistic for testing $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ with no constraints on $\boldsymbol{\mu}$ has the following form*

$$WST(\chi, \boldsymbol{\Sigma}_0) = \frac{n}{2} \text{tr}[(\mathbf{I}_p - \boldsymbol{\Sigma}_0 \hat{\boldsymbol{\Sigma}}^{-1})^2] \quad (2.4)$$

where $\chi = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and the test statistic $WST(\chi, \boldsymbol{\Sigma}_0)$ tends to a χ^2 -distribution with $\frac{p(p+1)}{2}$ degrees of freedom under H_0 when $n \rightarrow \infty$.

Proof. Because the parameter $\boldsymbol{\mu}$ and \mathbf{S} in the score vector are replaced by $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, If there is no constrain on $\boldsymbol{\mu}$. That means $U_1(\chi, \boldsymbol{\theta}) = \mathbf{0}$, consequently $I(\chi, \boldsymbol{\theta}) =: \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{n}{2}(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1}) \end{pmatrix}$. For a further step, $T_{\boldsymbol{\theta}}(\chi) = (\hat{\boldsymbol{\mu}}', \text{vec}'(\hat{\boldsymbol{\Sigma}}))'$ is the maximum likelihood estimator of the parameter $\boldsymbol{\theta}$, so we have

$$\begin{aligned} WST(\chi, \boldsymbol{\Sigma}_0) &= \frac{n}{2} \text{vec}'(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \hat{\boldsymbol{\Sigma}}^{-1}) \text{vec}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \\ &= \frac{n}{2} \text{vec}'(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \text{vec}(\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)\hat{\boldsymbol{\Sigma}}^{-1}) \\ &= \frac{n}{2} \text{tr}((\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)\hat{\boldsymbol{\Sigma}}^{-1}) \\ &= \frac{n}{2} \text{tr}[(\mathbf{I}_p - \boldsymbol{\Sigma}_0 \hat{\boldsymbol{\Sigma}}^{-1})^2] \end{aligned}$$

□

Corollary 2.1. *Wald's score test statistic for testing $H_0 : \boldsymbol{\Sigma} = \mathbf{I}_p$ with no constraints on $\boldsymbol{\mu}$ has the following form*

$$WST(\chi, \mathbf{I}_p) = \frac{n}{2} \text{tr}[(\mathbf{I}_p - \hat{\boldsymbol{\Sigma}}^{-1})^2]$$

where $\chi = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and the test statistic $WST(\chi, \mathbf{I}_p)$ tends to a χ^2 -distribution with $\frac{p(p+1)}{2}$ degrees of freedom under H_0 when $n \rightarrow \infty$.

Corollary 2.2. *Wald's score test statistic for testing $H_0 : \boldsymbol{\Sigma} = \gamma \mathbf{I}_p$ with no constraints on $\boldsymbol{\mu}$ has the following form*

$$WST(\chi, \gamma \mathbf{I}_p) = \frac{n}{2} \text{tr} \left[\left(\mathbf{I}_p - \frac{\text{tr}(\hat{\boldsymbol{\Sigma}})}{p} \hat{\boldsymbol{\Sigma}}^{-1} \right)^2 \right]$$

where $\chi = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\gamma > 0$ is an unknown parameter. Then the test statistic $WST(\chi, \gamma \mathbf{I}_p)$ tends to a χ^2 -distribution with $\frac{p(p+1)}{2} - 1$ degrees of freedom under H_0 when $n \rightarrow \infty$.

Proof. Replace the $\boldsymbol{\Sigma}_0$ by $\hat{\gamma} \mathbf{I}_p$ according to (2.4), where $\hat{\gamma} = \frac{\text{tr}(\hat{\boldsymbol{\Sigma}})}{p}$ is the maximum likelihood estimator of γ . \square

2.1. CLT for LSS of a high-dimensional sample covariance matrix

Following the above Proposition 2.1, the statistics of Wald's score test for the hypothesis (1.1) can be transformed into the trace of a matrix concerned with the sample covariance matrix, i.e. a function of the eigenvalues of the sample covariance matrix. It exactly meets the requirements in the CLT for LSS of large dimensional covariance matrices in Bai and Silverstein (2004), so we can modify the classical Wald's score tests by this limiting tool. In order to expand the usable range of the modified Wald's score tests, a quick survey of an enhanced version of the CLT for LSS of large dimensional covariance matrices is cited from Jiang (2015), which excludes the strict condition on the 4th moment for a wider usage. Before quoting, we first introduce some basic concepts and notations.

Suppose $(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ to be an i.i.d sample from some p -dimensional distribution with mean $\mathbf{0}_p$ and covariance matrix \mathbf{I}_p , where $\boldsymbol{\xi}_i = (\xi_{1i}, \xi_{2i}, \dots, \xi_{pi})'$. The corresponding sample covariance matrix is

$$\mathbf{S}_n = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_i \boldsymbol{\xi}_i^*. \quad (2.5)$$

where $(\cdot)^*$ is conjugate transpose. For simplicity, F^q, F^{q_n} denote the Marčenko-Pastur law of index q and q_n respectively, where $q_n = \frac{p}{n} \rightarrow q \in [0, +\infty)$. $F_n^{\mathbf{S}_n}$

represents the ESD (Empirical Spectral Distribution) of the matrix \mathbf{S}_n . Define the empirical process $G_n := \{G_n(f)\}$ indexed by \mathcal{A} ,

$$G_n(f) = p \cdot \int_{-\infty}^{+\infty} f(x) [F_n^{\mathbf{S}_n} - F^{q_n}] (dx), \quad f \in \mathcal{A}, \quad (2.6)$$

where \mathcal{U} is an open set of the complex plane including the supporting set of F^q and \mathcal{A} be the set of analytic functions $f : \mathcal{U} \mapsto \mathbb{C}$. Define

$$\kappa = \begin{cases} 2, & \text{if the } \boldsymbol{\xi} - \text{variables are real,} \\ 1, & \text{if the } \boldsymbol{\xi} - \text{variables are complex.} \end{cases}$$

Then an enhanced version of the CLT for LSS of large dimensional covariance matrices from Jiang (2015) (Lemma 2.1) is quoted as following:

Lemma 2.1. *Assume:*

$f_1, \dots, f_k \in \mathcal{A}$, $\{\xi_{ij}\}$ are i.i.d. random variables, such that $E\xi_{11} = 0$, $E|\xi_{11}|^2 = \kappa - 1$, $E|\xi_{11}|^4 < \infty$ and the $\{\xi_{ij}\}$ satisfy the condition

$$\frac{1}{np} \sum_{ij} E|\xi_{ij}|^4 I(|\xi_{ij}| \geq \sqrt{n}\eta) \rightarrow 0$$

for any fixed $\eta > 0$. Moreover, $\frac{p}{n} = q_n \rightarrow q \in [0, +\infty)$ as $n, p \rightarrow \infty$ and $E(\xi_{11}^4) = \beta + \kappa + 1$, where β is a constant.

Then the random vector $(G_n(f_1), \dots, G_n(f_k))$ forms a tight sequence by index n , and it weakly converges to a k -dimensional Gaussian vector with mean vector

$$\mu(f_j) = -\frac{\kappa - 1}{2\pi i} \oint f_j(z) \frac{q \underline{m}^3(z)(1 + \underline{m}(z))}{[(1 - q)\underline{m}^2(z) + 2\underline{m}(z) + 1]^2} dz \quad (2.7)$$

$$-\frac{\beta q}{2\pi i} \oint f_j(z) \frac{\underline{m}^3(z)}{(1 + \underline{m}(z))[(1 - q)\underline{m}^2(z) + 2\underline{m}(z) + 1]} dz, \quad (2.8)$$

and covariance function

$$v(f_j, f_\ell) = -\frac{\kappa}{4\pi^2} \oint \oint \frac{f_j(z_1) f_\ell(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} d\underline{m}(z_1) d\underline{m}(z_2) \quad (2.9)$$

$$-\frac{\beta q}{4\pi^2} \oint \oint \frac{f_j(z_1) f_\ell(z_2)}{(1 + \underline{m}(z_1))^2 (1 + \underline{m}(z_2))^2} d\underline{m}(z_1) d\underline{m}(z_2) \quad (2.10)$$

where $j, \ell \in \{1, \dots, k\}$, and $\underline{m}(z) \equiv m_{F^q}(z)$ is the Stieltjes Transform of $F^q \equiv (1 - q)I_{[0, \infty)} + qF^q$. The contours all contain the support of F^q and non overlapping in both (2.9) and (2.10).

3. The Proposed Testing Statistics

In this section, $\chi = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is set as an independent and identically distributed sample from a p dimensional random vector \mathbf{X} with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. To test on the structure of covariance matrix, we consider the hypothesis

$$H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \quad \text{v.s.} \quad H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0,$$

which covers the identity hypothesis test $H_0 : \boldsymbol{\Sigma} = \mathbf{I}_p$ and sphericity hypothesis test $H_0 : \boldsymbol{\Sigma} = \gamma \mathbf{I}_p$ as the special cases.

It has been well studied under the normal distribution assumption with the classical setting of fixed p , such as Anderson (2003), Nagao (1973) and John (1971) etc. Also, Wald's score test was given in Wald(1943). They all lost their effectiveness as the dimension p was much higher and performed even worse for the non-Gaussian variables. So we hope to correct the Wald's score test for the hypothesis (1.1) by using Lemma 2.1, which makes the correction applicable for large dimensional data and non-Gaussian assumption.

Set $\tilde{\boldsymbol{\xi}}_i = \boldsymbol{\Sigma}_0^{-\frac{1}{2}}(\mathbf{x}_i - \boldsymbol{\mu})$, then the array $\{\tilde{\boldsymbol{\xi}}_i\}_{i=1, \dots, n}$ contains p -dimensional standardized variables under H_0 . If the parameter $\boldsymbol{\mu}$ is unknown, the sample mean is used instead. Then the Lemma 2.1 should be applied by $n-1$ instead of n by Zheng et al. (2015). Therefore, we define $\tilde{\boldsymbol{\Sigma}} = \frac{n}{n-1} \hat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_0^{-1}$, then $\tilde{\boldsymbol{\Sigma}}$ has the same LSD with \mathbf{S}_{n-1} defined in (2.5) with n substituted by $n-1$. Let

$$\widetilde{\text{WST}}(\chi, \boldsymbol{\Sigma}_0) = \frac{n}{2} \text{tr}[(\mathbf{I}_p - \tilde{\boldsymbol{\Sigma}}^{-1})^2], \quad (3.1)$$

it is also natural to apply the Lemma 2.1 with $n-1$ instead of n to the modified Wald's score test statistic $\frac{2}{n} \widetilde{\text{WST}}(\chi, \boldsymbol{\Sigma}_0)$. Thus, the theorem of large dimensional Wald's score test is proposed as below:

Theorem 3.1. *Suppose that the conditions of Lemma 2.1 hold, for hypothesis test $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, $\widetilde{\text{WST}}(\chi, \boldsymbol{\Sigma}_0)$ is defined as (3.1), set $p/(n-1) = q_n \rightarrow q \in [0, 1)$, and $f(x) = (1 - \frac{1}{x})^2$. Then, under H_0 and when $n \rightarrow \infty$, the correction to Wald's score test statistics is*

$$CWST(\chi, \boldsymbol{\Sigma}_0) = v(f)^{-\frac{1}{2}} \left[\frac{2}{n} \widetilde{\text{WST}}(\chi, \boldsymbol{\Sigma}_0) - p \cdot F^{q_n}(f) - \mu(f) \right] \Rightarrow N(0, 1), \quad (3.2)$$

where F^{q_n} is the Marčenko-Pastur law of index q_n , and $F^{q_n}(f)$, $\mu(f)$ and $v(f)$ are calculated in (3.3), (3.5) and (3.6), respectively.

Proof. By the derivation (3.1), we have

$$\begin{aligned} \frac{2}{n} \widetilde{\text{WST}}(\chi, \Sigma_0) &= \text{tr}[(\mathbf{I}_p - \tilde{\Sigma}^{-1})^2] \\ &= \sum_{i=1}^p \left(1 - \frac{1}{\lambda_i^{\tilde{\Sigma}}}\right)^2 = p \cdot \int \left(1 - \frac{1}{x}\right)^2 dF_n^{\tilde{\Sigma}}(x) \\ &= p \cdot \int f(x) d\left(F_n^{\tilde{\Sigma}}(x) - F^{q_n}(x)\right) + p \cdot F^{q_n}(f), \end{aligned}$$

where $(\lambda_i^{\tilde{\Sigma}}), i = 1, \dots, p$ and $F_n^{\tilde{\Sigma}}$ are the eigenvalues and the ESD of the matrix $\tilde{\Sigma}$, respectively. $F^{q_n}(f)$ denotes the integral of the function $f(x)$ by the density corresponding to the Marčenko-Pastur law of index q_n , that is

$$F^{q_n}(f) = \int_{-\infty}^{\infty} f(x) dF^{q_n}(x) = 1 - \frac{2}{(1 - q_n)} + \frac{1}{(1 - q_n)^3}, \quad \text{if } 0 \leq q_n < 1, \quad (3.3)$$

which is calculated in the Appendix A.

As the definition in (2.6), we have

$$G_n(f) = p \cdot \int f(x) d\left(F_n^{\tilde{\Sigma}}(x) - F^{q_n}(x)\right) = \frac{2}{n} \widetilde{\text{WST}}(\chi, \Sigma_0) - p \cdot F^{q_n}(f). \quad (3.4)$$

By Lemma 2.1, $G_n(f)$ weakly converges to a Gaussian vector with the mean

$$\mu(f) = \frac{-(\kappa - 1)q(2q^2 - 5q - 1)}{(1 - q)^4} + \frac{\beta q(2q^2 - 3q - 1)}{(q - 1)^3} \quad (3.5)$$

and variance

$$v(f) = \frac{2\kappa q^2(2q^3 - 12q^2 + 18q + 1)}{(q - 1)^8} + \frac{4\beta q^3(2 - q)^2}{(q - 1)^6}. \quad (3.6)$$

which are calculated in the Appendix A. Then, by Lemma 2.1 and (3.4), we arrive at

$$\frac{2}{n} \widetilde{\text{WST}}(\chi, \Sigma_0) - p \cdot F^{q_n}(f) \Rightarrow N(\mu(f), v(f)),$$

Finally, we get

$$CWST(\chi, \Sigma_0) = v(f)^{-\frac{1}{2}} \left[\frac{2}{n} \widetilde{\text{WST}}(\chi, \Sigma_0) - p \cdot F^{q_n}(f) - \mu(f) \right] \Rightarrow N(0, 1)$$

□

For the identity and sphericity hypothesis tests, the corollaries are as below :

Corollary 3.1. *For testing $H_0 : \Sigma = \mathbf{I}_p$ with no constraints on $\boldsymbol{\mu}$, the conclusion of Theorem 3.1 still holds, only with the test statistic $\widetilde{WST}(\chi, \mathbf{I}_p)$ in (3.2) is revised by*

$$\widetilde{WST}(\chi, \mathbf{I}_p) = \frac{n}{2} \text{tr}[(\mathbf{I}_p - (\frac{n}{n-1} \widehat{\Sigma})^{-1})^2]. \quad (3.7)$$

Corollary 3.2. *For testing $H_0 : \Sigma = \gamma \mathbf{I}_p$ with no constraints on $\boldsymbol{\mu}$, the conclusion of Theorem 3.1 still holds, only with the test statistic $\widetilde{WST}(\chi, \mathbf{I}_p)$ in (3.2) is revised by*

$$\widetilde{WST}(\chi, \gamma \mathbf{I}_p) = \frac{n}{2} \text{tr}[(\mathbf{I}_p - \hat{\gamma}(\frac{n}{n-1} \widehat{\Sigma})^{-1})^2]$$

where $\hat{\gamma} = \frac{\text{tr}(\frac{n}{n-1} \widehat{\Sigma})}{p}$ is the maximum likelihood estimator of γ .

4. Simulation Study

Without loss of the generality, the identity hypothesis test $H_0 : \Sigma = \mathbf{I}_p$ is investigated in this section. The simulations are conducted to compare our proposed corrections to Wald's score test (CWST) with other large dimensional tests on covariance matrices, like the tests in Ledoit and Wolf (2002) (LWT), Cai and Ma (2013) (CMT) and classical covariance tests in Nagao (1973) (NHT) and Wald (1943) (WST). We generate i.i.d random samples $\chi = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ from two scenarios of the p -dimensional populations under the null hypothesis:

Normal Assumption: Following a p -dimensional normal distribution with mean $\mu_0 \mathbf{1}_p$ and covariance matrix \mathbf{I}_p , where $\mu_0 = 2$ and $\mathbf{1}_p$ denotes a vector with that all elements are 1.

Gamma Assumption: Following a p -dimensional Gamma distribution with all the components are i.i.d. from the distribution of Gamma (4,0.5).

For each scenario, The empirical sizes and powers are reported with 10,000 replications at $\alpha = 0.05$ significance level. We chose the cases from $n = 300, p = 80, 120, 160, 200$ and $n = 500, p = 80, 160, 240, 320$ and the mean

parameter is supposed to be unknown and substituted by the sample mean during the calculations.

For the alternative hypothesis, the population covariance matrix is designed as the tridiagonal matrix $\Sigma = (\sigma_{ij})$ for different population assumptions, where for $\rho \in (0, 1)$ and

$$\sigma_{ij} = \begin{cases} 1, & i = j \\ \rho, & |i - j| = 1 \\ 0, & |i - j| > 1 \end{cases} \quad (4.1)$$

Simulation results of empirical sizes and powers of four tests are listed in the Table 1, which includes our proposed CWST, the test in Ledoit and Wolf (2002) (LWT) and Nagao (1973) (NHT) and Wald's score test (WST). The comparison of empirical sizes between our proposed CWST and the test in Cai and Ma (2013) (CMT) is also presented in Table 2.

Note from the Table 1, it is easily to find out the informations as below:

- (i) The traditional Wald's score test (WST) is completely unworkable for large dimensional data in respect of their empirical sizes, which all equals to 1 for any case listed in the table.
- (ii) The empirical sizes of the test in Nagao(1973) (NHT) also deviate far from the nominal test size 5%, and they increase with the dimension p , especially worse for Gamma Assumption.
- (iii) For the cases of the relatively smaller dimensions in the table, like $p = 80$ with $n = 300$ or $p = 160$ with $n = 500$, the test in Ledoit and wolf (2002) (LWT) behaves well and provides higher powers. However, the empirical sizes of LWT rise up against the nominal level as p increases, and it shows a even worse result under the Gamma Assumption, where our proposed CWST is still active.
- (iv) The empirical sizes of our proposed test CWST are almost around the nominal size 5% for both distribution assumptions. Although the empirical powers of our proposed CWST are not as higher as others' for the very small ρ , the other tests work well in one place but fail in another because of their rising empirical sizes. Besides, the empirical powers of the proposed CWST quickly increase to 1 with a slight upward adjustment of ρ .

Table 1: Empirical sizes and powers of the comparative tests for $H_0 : \mathbf{\Sigma} = \mathbf{I}_p$ at $\alpha = 0.05$ significance level for Normal and Gamma Assumptions with 10,000 replications. The alternative hypothesis is the tridiagonal matrix $\mathbf{\Sigma} = (\sigma_{ij})$ with σ_{ij} defined in (4.1).

	$(n, p) = (300, 80)$					
	Sizes	Powers		Sizes	Powers	
	$\rho = 0$	$\rho = 0.05$	$\rho = 0.15$	$\rho = 0$	$\rho = 0.05$	$\rho = 0.15$
		Normal			Gamma	
CWST	0.0649	0.1303	0.9861	0.0640	0.1273	0.9748
LWT	0.0704	0.2199	1	0.2290	0.4698	1
NHT	0.1191	0.2632	0.9998	0.2934	0.4701	0.9998
WST	1	1	1	1	1	1
	$(n, p) = (300, 160)$					
	Sizes	Powers		Sizes	Powers	
	$\rho = 0$	$\rho = 0.05$	$\rho = 0.18$	$\rho = 0$	$\rho = 0.05$	$\rho = 0.18$
		Normal			Gamma	
CWST	0.0682	0.1071	0.9785	0.0688	0.1033	0.9746
LWT	0.0874	0.2619	1	0.2607	0.5358	1
NHT	0.1779	0.3213	1	0.3486	0.5182	1
WST	1	1	1	1	1	1
	$(n, p) = (500, 160)$					
	Sizes	Powers		Sizes	Powers	
	$\rho = 0$	$\rho = 0.05$	$\rho = 0.12$	$\rho = 0$	$\rho = 0.05$	$\rho = 0.12$
		Normal			Gamma	
CWST	0.0615	0.1713	0.9820	0.0596	0.1574	0.9725
LWT	0.0694	0.3982	1	0.2279	0.6774	1
NHT	0.1276	0.4152	1	0.3067	0.6321	1
WST	1	1	1	1	1	1
	$(n, p) = (500, 320)$					
	Sizes	Powers		Sizes	Powers	
	$\rho = 0$	$\rho = 0.05$	$\rho = 0.15$	$\rho = 0$	$\rho = 0.05$	$\rho = 0.15$
		Normal			Gamma	
CWST	0.0627	0.1139	0.9349	0.0640	0.1127	0.9304
LWT	0.0912	0.4669	1	0.2887	0.7368	1
NHT	0.1904	0.4750	1	0.3857	0.6341	1
WST	1	1	1	1	1	1

Furthermore, it was showed in Cai and Ma (2013) that CMT provided the optimal powers, which uniformly dominated that of the corrected LRTs by random matrix theory over the asymptotic regime. However, in contrast to CMT, the proposed CWST performs more steady empirical sizes, which is supported by the following brief table 2. The corresponding power comparisons for both scenarios are similar to the ones between LWT and CWST under the Normal Assumption. It shows a relatively slow ascent of the powers by our proposed test due to the involvement of the inverse of the sample covariance matrix. But the powers of our test will quickly rise to 1 if the null and alternative hypotheses are not much close. And if it will bring serious consequences when the null hypothesis occurred but not found in the practice, we should be strict to Type I errors. Therefore, it is better to choose our proposed test in such a situation, which provides precise empirical sizes.

Table 2: Empirical sizes of the tests proposed in this paper(CWST) and Cai and Ma (2013) (CMT) for $H_0 : \Sigma = \mathbf{I}_p$ at $\alpha = 0.05$ significance level. The simulations are repeated 10,000 times for Normal and Gamma Assumptions respectively.

$n = 300$	CWST	CMT	CWST	CMT
	Normal		Gamma	
$p = 80$	0.0649	0.0710	0.0640	0.0772
$p = 120$	0.0652	0.0789	0.0687	0.0795
$p = 160$	0.0682	0.0872	0.0688	0.0905
$p = 200$	0.0719	0.0947	0.0713	0.0973
$n = 500$	CWST	CMT	CWST	CMT
	Normal		Gamma	
$p = 80$	0.0587	0.0620	0.0584	0.0603
$p = 160$	0.0615	0.0699	0.0596	0.0678
$p = 240$	0.0623	0.0855	0.0603	0.0832
$p = 320$	0.0627	0.0914	0.0640	0.0953

5. Conclusion

In this paper, the new tests for the covariance matrices structure based on modification of Wald's score test are proposed by RMT. They are feasible

for large dimensional data without restriction of population distribution and provide the accurate empirical sizes. However, it must be noted that the proposed CWST cannot be used for $p \geq n$, even if the case of $p/n \rightarrow 1$, since it is involved with the inverse of the sample covariance matrix. That's also the reason why it gives a relatively slow ascent of the powers. This problem is brought by the statistic on which our correction is based, rather than the the idea of correction or the large dimensional CLT in random matrix theory we used. So it can be expected that the similar idea of methods can be used for other suitable statistics to avoid these disadvantages. For example, Jiang(2015) proposed a testing statistic for the large dimensional covariance structure test based on amending Rao's score tests, which is applicable for the case of $p > n$ and Non-Gaussian assumption. In the future, we may also look forward to the applications of random matrix theory in more statistical inferences.

Acknowledgement

The author thanks the reviewers for their helpful comments and suggestions to make an improvement of this article. This research was supported by the National Natural Science Foundation of China 11471140.

Appendix A. Proofs

• Calculation of $F^{q_n}(f)$ in (3.3).

Because

$$F^{q_n}(f) = \int_{-\infty}^{\infty} f(x)(d)F^{q_n}(x) = \int_{-\infty}^{\infty} \frac{(1 - \frac{1}{x})^2}{2\pi x q_n} \sqrt{(b_n - x)(x - a_n)} dx$$

Use the substitution $x = 1 + q_n - 2\sqrt{q_n} \cos \theta = -2\sqrt{q_n}(\cos \theta + d_0)$,

where $0 \leq \theta \leq \pi$ and $d_0 = -\frac{1+q_n}{2\sqrt{q_n}}$ is a constant. Then

$$\begin{aligned}
F^{q_n}(f) &= \int_{a_n}^{b_n} \frac{(1 - \frac{1}{x})^2}{2\pi x q_n} \sqrt{(b_n - x)(x - a_n)} dx \\
&= \frac{2}{\pi} \int_0^\pi \frac{(q_n - 2\sqrt{q_n} \cos \theta)^2}{(1 + q_n - 2\sqrt{q_n} \cos \theta)^3} \sqrt{\sin^2 \theta} \sin \theta d\theta \\
&= \frac{1}{\pi} \int_0^{2\pi} \frac{[-2\sqrt{q_n}(\cos \theta + d_0) - 1]^2 \sin^2 \theta}{[-2\sqrt{q_n}(\cos \theta + d_0)]^3} d\theta \\
&= \frac{1}{\pi} \int_0^{2\pi} \left[\frac{\sin^2 \theta}{-2\sqrt{q_n}(\cos \theta + d_0)} - \frac{2 \sin^2 \theta}{4q(\cos \theta + d_0)^2} - \frac{\sin^2 \theta}{8q_n^{\frac{3}{2}}(\cos \theta + d_0)^3} \right] d\theta
\end{aligned}$$

First, we have

$$\begin{aligned}
\int_0^{2\pi} \frac{1}{\cos \theta + d_0} d\theta &= \int_0^{2\pi} \frac{1}{\cos^2 \frac{\theta}{2} - \sin^2 \frac{\theta}{2} + d_0} d\theta \\
&= \int_0^{2\pi} \frac{2d\frac{\theta}{2}}{(1 - \tan^2 \frac{\theta}{2} + d_0 \sec^2 \frac{\theta}{2}) \cos^2 \frac{\theta}{2}} \\
&= \int_0^{2\pi} \frac{2d \tan \frac{\theta}{2}}{d_0 + 1 + (d_0 - 1) \tan^2 \frac{\theta}{2}} \\
&= \frac{2}{\sqrt{d_0^2 - 1}} \int_0^{2\pi} \frac{1}{1 + \left(\sqrt{\frac{d_0 - 1}{d_0 + 1}} \tan \frac{\theta}{2} \right)^2} d \left(\sqrt{\frac{d_0 - 1}{d_0 + 1}} \tan \frac{\theta}{2} \right) \\
&= \frac{-2\pi}{\sqrt{d_0^2 - 1}} \tag{A.1}
\end{aligned}$$

So the first part of the integral,

$$\begin{aligned}
&\int_0^{2\pi} -\frac{\sin^2 \theta}{2\sqrt{q_n}(\cos \theta + d_0)} d\theta \\
&= -\frac{1}{2\sqrt{q_n}} \int_0^{2\pi} \frac{1 - \cos^2 \theta}{\cos \theta + d_0} d\theta \\
&= -\frac{1}{2\sqrt{q_n}} \int_0^{2\pi} \left(-\cos \theta + d_0 + \frac{1 - d_0^2}{\cos \theta + d_0} \right) d\theta
\end{aligned}$$

$$= \frac{\pi}{2q_n}(1 + q_n - |1 - q_n|) \quad (\text{A.2})$$

Secondly, the middle part of the integral can be calculated by the equation eq.(A.1), which is

$$\begin{aligned} \int_0^{2\pi} -\frac{2\sin^2\theta}{4q_n(\cos\theta + d_0)^2}d\theta &= \frac{1}{2q_n} \int_0^{2\pi} \left(\frac{-1}{\cos\theta + d_0} \right)' \sin\theta d\theta \\ &= \frac{1}{2q_n} \int_0^{2\pi} \frac{\cos\theta}{\cos\theta + d_0} d\theta = \frac{\pi}{q_n} \left(1 - \frac{1 + q_n}{|1 - q_n|} \right). \end{aligned} \quad (\text{A.3})$$

For the third part of the limiting integral $F^{q_n}(f)$, the following integral is calculated first.

$$\begin{aligned} &\int_0^{2\pi} \frac{1}{(\cos\theta + d_0)^2} d\theta \\ &= \int_0^{2\pi} \frac{\sin^2\theta}{(\cos\theta + d_0)^2} d\theta + \int_0^{2\pi} \frac{\cos\theta - d_0}{\cos\theta + d_0} d\theta + \int_0^{2\pi} \frac{d_0^2}{(\cos\theta + d_0)^2} d\theta \end{aligned}$$

Then, by the eq. (A.1) and (A.3), we arrive at

$$\begin{aligned} &\int_0^{2\pi} \frac{1}{(\cos\theta + d_0)^2} d\theta \\ &= \frac{1}{d_0^2 - 1} \left[-\int_0^{2\pi} \frac{\sin^2\theta}{(\cos\theta + d_0)^2} d\theta - \int_0^{2\pi} \frac{\cos\theta - d_0}{\cos\theta + d_0} d\theta \right] \\ &= \frac{8\pi q_n(1 + q_n)}{|1 - q_n|^3} \end{aligned} \quad (\text{A.4})$$

So the third part of the limiting integral $F^{q_n}(f)$ is

$$\begin{aligned} &\int_0^{2\pi} -\frac{\sin^2\theta}{8q_n^{\frac{3}{2}}(\cos\theta + d_0)^3} d\theta \\ &= \frac{1}{16q_n^{\frac{3}{2}}} \int_0^{2\pi} \left[\frac{-1}{(\cos\theta + d_0)^2} \right]' \sin\theta d\theta = \frac{1}{16q_n^{\frac{3}{2}}} \int_0^{2\pi} \frac{\cos\theta}{(\cos\theta + d_0)^2} d\theta \\ &= \frac{1}{16q_n^{\frac{3}{2}}} \left[\int_0^{2\pi} \frac{1}{\cos\theta + d_0} d\theta - d_0 \int_0^{2\pi} \frac{1}{(\cos\theta + d_0)^2} d\theta \right] = \frac{\pi}{|1 - q_n|^3} \end{aligned}$$

Above all, $F^{q_n}(f)$ is presented below

$$\begin{aligned} F^{q_n}(f) &= \frac{1}{\pi} \left[\frac{\pi}{2q_n} (1 + q_n - |1 - q_n|) + \frac{\pi}{q_n} \left(1 - \frac{1 + q_n}{|1 - q_n|} \right) + \frac{\pi}{|1 - q_n|^3} \right] \\ &= \begin{cases} 1 - \frac{2}{(1 - q_n)} + \frac{1}{(1 - q_n)^3}, & \text{if } 0 \leq q_n < 1, \\ \frac{1}{q_n} - \frac{2}{q_n(q_n - 1)} + \frac{1}{(q_n - 1)^3}, & \text{if } q_n > 1. \end{cases} \end{aligned}$$

It is worth to note that the $F^{q_n}(f)$ should plus the term $(1 - 1/x)^2 \cdot (1 - \frac{1}{q_n})$ at the origin $x = 0$ if $q_n > 1$, which is obviously infinity. That's one of reasons that the correction to Wald's score test can only be used as $q_n < 1$. Then we arrive at

$$F^{q_n}(f) = 1 - \frac{2}{(1 - q_n)} + \frac{1}{(1 - q_n)^3}, \quad \text{if } 0 \leq q_n < 1.$$

• **Calculation of $\mu(f)$ in (3.5).**

In the same way, with $H(t) = I_{[1, \infty)}(t)$, the first part of $\mu(f)$ is also obtained by (9.12.13) in Bai and Silverstein (2010),

$$\mu_1(f) = (\kappa - 1) \cdot \left(\frac{f(a(q)) + f(b(q))}{4} - \frac{1}{2\pi} \int_{a(q)}^{b(q)} \frac{f(x)}{\sqrt{4q - (x - 1 - q)^2}} dx \right)$$

where $a(q) = (1 - \sqrt{q})^2$ and $b(q) = (1 + \sqrt{q})^2$. For $f(x) = (1 - \frac{1}{x})^2$, make a substitution $x = 1 + q - 2\sqrt{q} \cos \theta$, $0 \leq \theta \leq \pi$, then

$$\begin{aligned} \mu_1(f) &= (\kappa - 1) \left(\frac{f(a(q)) + f(b(q))}{4} - \frac{1}{4\pi} \int_0^{2\pi} f(1 + q - 2\sqrt{q} \cos \theta) d\theta \right) \\ &= (\kappa - 1) \left(\frac{q^4 - 6q^3 + 9q^2 + 4q}{2(1 - q)^4} \right. \\ &\quad \left. - \frac{1}{4\pi} \int_0^{2\pi} \left[1 - \frac{2}{-2\sqrt{q}(\cos \theta + d_0^*)} + \frac{1}{4q(\cos \theta + d_0^*)^2} \right] d\theta \right) \\ &= (\kappa - 1) \left(\frac{q^4 - 6q^3 + 9q^2 + 4q}{2(1 - q)^4} - \frac{1}{2} + \frac{1}{|1 - q|} - \frac{1 + q}{2|1 - q|^3} \right) \end{aligned}$$

where $d_0^* = -\frac{1+q}{2\sqrt{q}}$ is an analogy to the constant d_0 with q instead of q_n , and the calculation can also be based on (A.1) and (A.4). Because the correction to Wald's score test can only be applied to the case $q_n < 1$, so we only choose the first case

$$\mu_1(f) = \frac{-(\kappa-1)q(2q^2-5q-1)}{(1-q)^4}, \quad \text{if } 0 \leq q < 1.$$

For the second part of $\mu(f)$, by (2.8) we have

$$\mu_2(f) = -\frac{\beta q}{2\pi i} \oint \left(1 - \frac{1}{z}\right)^2 \frac{\underline{m}^3(z)}{(1 + \underline{m}(z))[(1-q)\underline{m}^2(z) + 2\underline{m}(z) + 1]} dz,$$

Recall the equation (9.12.12) in Bai and Silverstein (2010)

$$z = -\frac{1}{\underline{m}(z)} + \frac{q}{1 + \underline{m}(z)}, \quad (\text{A.5})$$

it is easily obtained that

$$\left(1 - \frac{1}{z}\right)^2 = \frac{[m^2 - (q-2)m + 1]^2}{[(q-1)m - 1]^2} \quad dz = \frac{(1-q)m^2 + 2m + 1}{m^2(1+m)^2} dm$$

where $\underline{m}(z)$ is denoted as m for simplicity if no confusion. Then we have

$$\mu_2(f) = -\frac{\beta q}{2\pi i(q-1)^2} \oint \frac{m[m^2 - (q-2)m + 1]^2}{(m - \frac{1}{q-1})^2(1+m)^3} dm,$$

and the contour for the integral of m is obtained by solving the equation (A.5), which should enclose the interval $\left[-\frac{1}{1-\sqrt{q}}, -\frac{1}{1+\sqrt{q}}\right]$ when $0 \leq q < 1$. Therefore, -1 and $\frac{1}{q-1}$ are the residues if $q < 1$, and the integral is calculated as

$$\mu_2(f) = \frac{\beta q(2q^2 - 3q - 1)}{(q-1)^3}.$$

Finally, we obtained

$$\mu(f) = \frac{-(\kappa-1)q(2q^2-5q-1)}{(1-q)^4} + \frac{\beta q(2q^2-3q-1)}{(q-1)^3}.$$

• **Calculation of $v(f)$.**

By Lemma 2.1, we have

$$\begin{aligned} v(f_j, f_\ell) &= -\frac{\kappa}{4\pi^2} \oint \oint \frac{f_j(z_1)f_\ell(z_2)}{(\underline{m}(z_1) - \underline{m}(z_2))^2} d\underline{m}(z_1) d\underline{m}(z_2) \\ &\quad - \frac{\beta q}{4\pi^2} \oint \oint \frac{f_j(z_1)f_\ell(z_2)}{(1 + \underline{m}(z_1))^2(1 + \underline{m}(z_2))^2} d\underline{m}(z_1) d\underline{m}(z_2), \end{aligned}$$

and

$$\begin{aligned} f(z_1)f(z_2) &= \left(1 - \frac{1}{z_1}\right)^2 \left(1 - \frac{1}{z_2}\right)^2 \\ &= 1 - \frac{2}{z_1} - \frac{2}{z_2} + \frac{1}{z_1^2} + \frac{1}{z_2^2} + \frac{4}{z_1 z_2} - \frac{2}{z_1^2 z_2} - \frac{2}{z_1 z_2^2} + \frac{1}{z_1^2 z_2^2}. \end{aligned}$$

It is known that $v(\mathbf{1}, \mathbf{1}) = 0$, where $\mathbf{1}$ denote constant function which equals to 1. For $z \in \mathbb{C}^+$, the following equation is given in Bai and Silverstein (2010),

$$z = -\frac{1}{\underline{m}(z)} + \frac{q}{1 + \underline{m}(z)}. \quad (\text{A.6})$$

Still use m_i to simplify $\underline{m}(z_i)$, $i = 1, 2$. For fixed m_2 , there is a contour enclosed $(q-1)^{-1}$ and -1 as poles when $0 \leq q < 1$. Then, for the first item of $v(f)$, we have

$$\begin{aligned} &\oint \frac{1}{z_1} \cdot \frac{1}{(m_1 - m_2)^2} dm_1 \\ &= \oint \frac{m_1(1 + m_1)}{(q-1)(m_1 - \frac{1}{q-1})(m_1 - m_2)^2} dm_1 \\ &= 2\pi i \cdot \frac{q}{(q-1)^3(m_2 - \frac{1}{q-1})^2}. \end{aligned}$$

and

$$\begin{aligned} &\oint \frac{1}{z_1^2} \cdot \frac{1}{(m_1 - m_2)^2} dm_1 \\ &= \oint \frac{m_1^2(1 + m_1)^2}{(q-1)^2(m_1 - \frac{1}{q-1})^2} (m_2 - \frac{1}{q-1})^{-2} \left(1 - \frac{m_1 - \frac{1}{q-1}}{m_2 - \frac{1}{q-1}}\right)^{-2} dm_1 \\ &= 4\pi i \left[\frac{q(1+q)}{(q-1)^5(m_2 - \frac{1}{q-1})^2} + \frac{q^2}{(q-1)^6(m_2 - \frac{1}{q-1})^3} \right]. \quad (\text{A.7}) \end{aligned}$$

So $v(\frac{1}{z_1^2} - \frac{2}{z_1}, \mathbf{1}) = 0$. Similarly, $v(\mathbf{1}, \frac{1}{z_1^2} - \frac{2}{z_1}) = 0$.

Therefore, there are only four parts left, i.e. $\frac{4}{z_1 z_2} - \frac{2}{z_1^2 z_2} - \frac{2}{z_1 z_2^2} + \frac{1}{z_1^2 z_2^2}$. Further,

$$\begin{aligned} v(\frac{1}{z_1}, \frac{1}{z_2}) &= -\frac{\kappa}{4\pi^2} \oint \frac{1}{z_2} \oint \frac{1}{z_1} \frac{1}{(m_1 - m_2)^2} dm_1 dm_2 \\ &= \frac{\kappa q}{2\pi i (q-1)^3} \oint \frac{m_2(1+m_2)}{(q-1)(m_2 - \frac{1}{q-1})^3} dm_2 \\ &= \frac{\kappa q}{(q-1)^4} \end{aligned}$$

$$\begin{aligned} v(\frac{1}{z_1}, \frac{1}{z_2^2}) &= -\frac{\kappa}{4\pi^2} \oint \frac{1}{z_2^2} \oint \frac{1}{z_1} \frac{1}{(m_1 - m_2)^2} dm_1 dm_2 \\ &= \frac{\kappa q}{2\pi i (q-1)^5} \oint \frac{m_2^2(1+m_2)^2}{(m_2 - \frac{1}{q-1})^4} dm_2 \\ &= \frac{2\kappa q(1+q)}{(q-1)^6}. \end{aligned}$$

Similarly, $v(\frac{1}{z_1^2}, \frac{1}{z_2}) = \frac{2\kappa q(1+q)}{(q-1)^6}$. For the last part $v(\frac{1}{z_1^2}, \frac{1}{z_2^2})$, the integral is calculated by eq.(A.7) as below.

$$\begin{aligned} v(\frac{1}{z_1^2}, \frac{1}{z_2^2}) &= -\frac{\kappa}{4\pi^2} \oint \frac{1}{z_2^2} \oint \frac{1}{z_1^2} \frac{1}{(m_1 - m_2)^2} dm_1 dm_2 \\ &= \frac{\kappa}{\pi i} \oint \frac{m_2^2(1+m_2)^2}{(q-1)^2(m_2 - \frac{1}{q-1})^2} \left[\frac{q(1+q)}{(q-1)^5(m_2 - \frac{1}{q-1})^2} + \frac{q^2}{(q-1)^6(m_2 - \frac{1}{q-1})^3} \right] dm_2 \\ &= \frac{2\kappa q(1+2q)(q+2)}{(q-1)^8} \end{aligned}$$

So the first item of $v(f)$ is

$$\begin{aligned} v_1(f) &= 4v(\frac{1}{z_1}, \frac{1}{z_2}) - 2v(\frac{1}{z_1^2}, \frac{1}{z_2}) - 2v(\frac{1}{z_1}, \frac{1}{z_2^2}) + v(\frac{1}{z_1^2}, \frac{1}{z_2^2}) \\ &= \frac{2\kappa q^2(2q^3 - 12q^2 + 18q + 1)}{(q-1)^8} \end{aligned}$$

when $0 \leq q < 1$.

Secondly, the latter item of $v(f)$ is

$$v_2(f) = -\frac{\beta q}{4\pi^2} \oint \oint \frac{f(z_1)f(z_2)}{(1+m_1)^2(1+m_2)^2} dm_1 dm_2.$$

Furthermore,

$$\begin{aligned} \oint \frac{f(z_1)}{(1+m_1)^2} dm_1 &= \oint \frac{[m_1^2 - (q-2)m_1 + 1]^2}{[(q-1)m_1 - 1]^2(1+m_1)^2} dm_1 \\ &= 4\pi i \frac{q(2-q)}{(q-1)^3} \end{aligned}$$

since the contour contains $\frac{1}{q-1}$ and -1 as residues if $0 \leq q \leq 1$. Thus we get

$$v_2(f) = -\frac{\beta q}{4\pi^2} \cdot \left(4\pi i \frac{q(2-q)}{(q-1)^3} \right)^2 = \frac{4\beta q^3(2-q)^2}{(q-1)^6}$$

Finally, we obtained

$$v(f) = \frac{2\kappa q^2(2q^3 - 12q^2 + 18q + 1)}{(q-1)^8} + \frac{4\beta q^3(2-q)^2}{(q-1)^6}.$$

References

- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Third edition. Wiley, New York.
- Bai, Z.D., Jiang, D.D., Yao, J. F. and Zheng, S. (2009). Corrections to LRT on large dimensional covariance matrix by RMT. *Ann. Statist.*, 37, No.6B: 3822-3840.
- Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. *Ann.Probab.*, 32: 553-605.
- Bai, Z. D. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*. 2nd ed., Beijing: Science Press.
- Cai, T.T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, 19, 5B: 2359-2388.

- Chen, S.X., Zhang, L.X. and Zhong, P.S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.*, 105: 810-819.
- Gombay, E. (2002). Parametric sequential tests in the presence of nuisance parameters. *Theory Stochastic Progresses*, 8: 106-118.
- Jiang, D. (2015). Tests for large dimensional covariance structure based on Rao's score test. *Submitted* arXiv:1510.03098v2
- Jiang, D., Jiang, T. and Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference*, 142: 2241-2256.
- John, S. (1971). Some Optimal Multivariate Tests. *Biometrika*, 59: 123-127.
- Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29: 295-327.
- Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, *Ann. Statist.*, 30: 1081-1102.
- Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, 1: 700-709.
- Rao, C.R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 44: 50-57.
- Srivastava, M.S. (2005). Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, 35: 251-272.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.*, 54: 426-482.
- Zheng, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F -Matrices. *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, **48**, No. 2: 444-476.
- Zheng, S., Bai, Z.D. and Yao, J. F. (2015). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.*, 43, No. 2: 546-591.